STT200 2-18-09a

SAYS $\sqrt{\frac{n}{n-1}} \sim 1$

$s_x \doteq \hat{\sigma}_x$

QUIZ 2 2-17-09 (one version)

$\bar{x} = 6 \quad \bar{y} = 10 \quad s_x = 3 \quad s_y = 5 \quad r = 0.7 \quad n = 300$

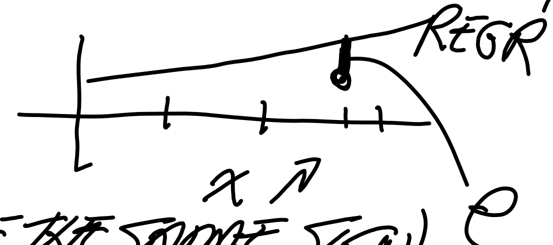
1. SLOPE NAIVE = $s_y / s_x = 5/3 \cong 1.67$

2. SLOPE REGR = $r \, s_y / s_x = .7(5/3) = 3.5/3 \cong 1.17$

NOTE: r = SLOPE REGR / SLOPE NAIVE

3. FRACTION OF s_y^2 EXPLAINED BY REGRESSION $= r^2 = .7^2 = .49$

NOTE: $r^2 = 1 - s_e^2 / s_y^2 = (s_y^2 - s_e^2)/s_y^2$



4. $r[6x+2, 3y-4] = r[x, y]$ (6 & 3 HAVE THE SAME SIGN)

5. PREDICTED y FOR $x = 9$ IS $\hat{y} = 10$ SINCE $(\bar{x}, \bar{y}) = (6, 10)$ LIES ON REGR LINE

SO FROM $x = 6$ PREDICT $y = 10$ BUT $9 = \bar{x} + s_x$ PRED $10 + .7(5)$ ANS.

FORMULA PREDICT $\bar{y} + (x - \bar{x}) \, r \, s_y / s_x$

QUIZ CONT. $\bar{x}=6$ $\bar{y}=10$ $s_x=3$ $s_y=5$ $r=0.7$ $n=300$

6. PREDICTED y FOR $x=12$ IS $\hat{y}=\bar{y}+(12-\bar{x})\,\text{SLOPE}$

$\hat{y}=10+(12-6)\,1.17..=17$

7. ESTIMATE OF μ_y IF WE KNOW $\mu_x=6$ IS 10.

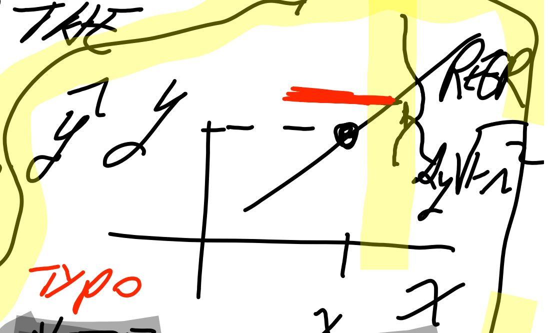
BY ALGEBRA: REGRESS OF $\mu_y=\underset{10}{\bar{y}}+\underset{6-6=0}{(\mu_x-\bar{x})}\,r\,s_y/s_x$.

8. STANDARD DEVIATION OF ALL y WITH $x=19$ IS (FOR ELLIPTICAL PLOTS) $\sqrt{1-r^2}\,s_y=\sqrt{1-.7^2}\,5\cong 3.57$

9. REGRESSION OF y ON x (USUAL) IS NOT THE SAME AS REGRESSION OF x ON y.

10. $r[x,y]=\dfrac{\overline{xy}-\bar{x}\,\bar{y}}{\sqrt{\overline{x^2}-\bar{x}^2}\sqrt{\overline{y^2}-\bar{y}^2}}=r[y,x]$

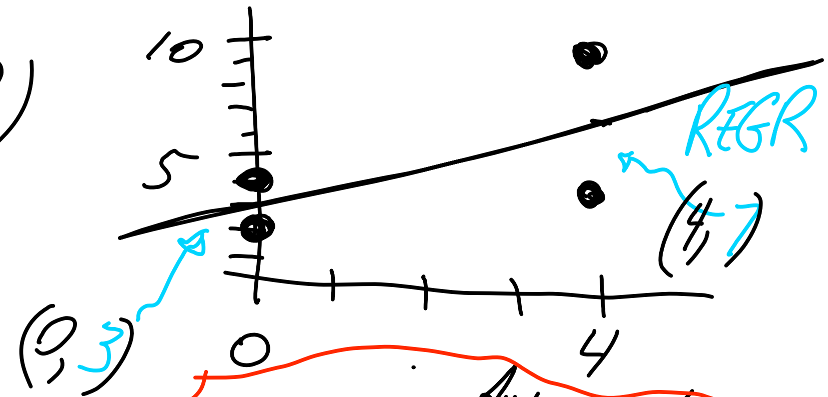
#8



TYPO

NOTE $r[x,\hat{y}]=|r|$

DATA (0,2) (0,4) (4,4) (4,10)



11. $\text{Slope} = \frac{\text{RISE}}{\text{RUN}} = \frac{7-3}{4-0} = 4/4 = 1$

12. y-INTERCEPT $= b_0 = 3$

$\text{Slope} = r\, s_y/s_x = b_1$

$b_0 = \bar{y} + (0-\bar{x})\,\text{slope}$

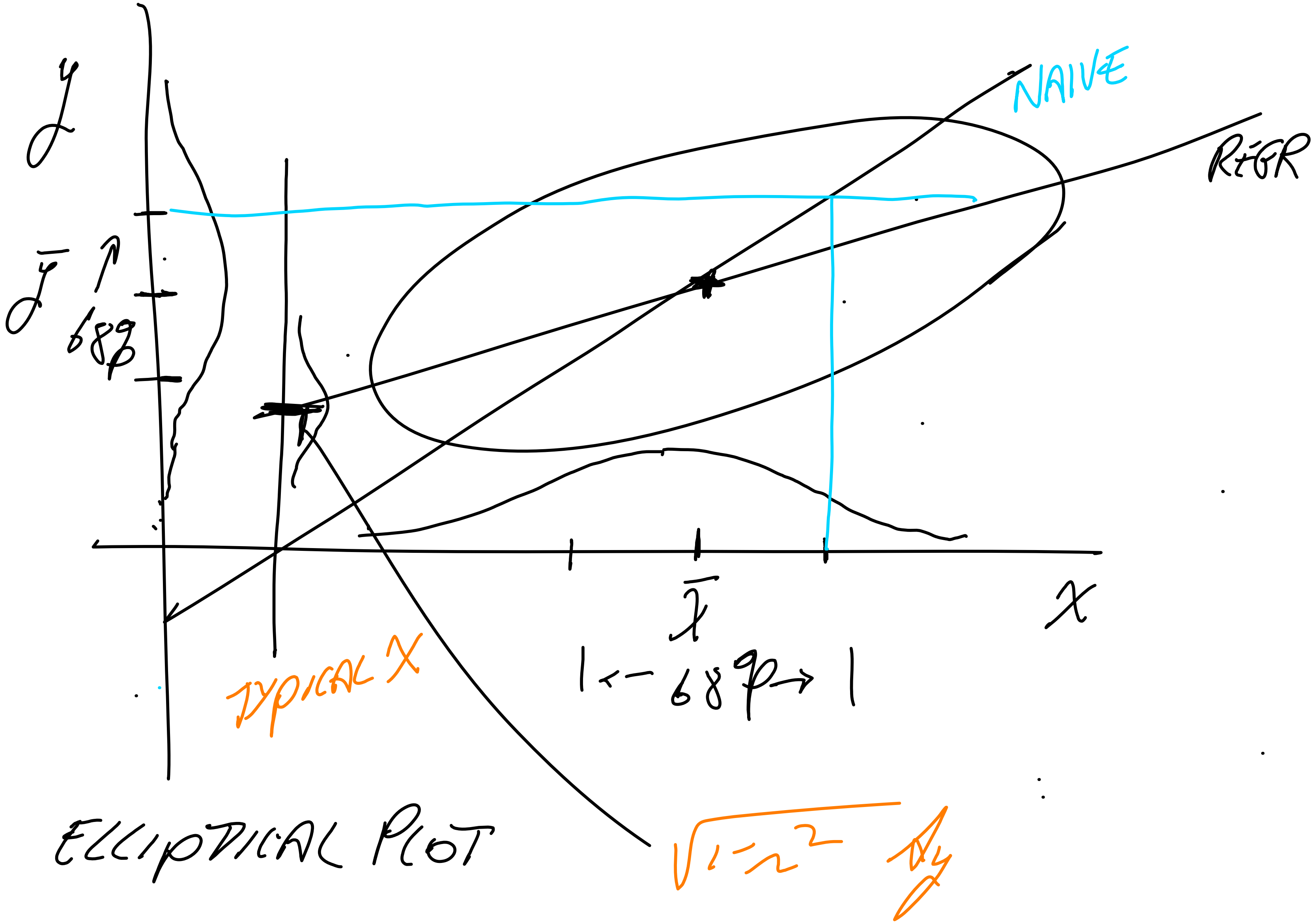
DATA $\{2, 4, 5\}$

13. $s_x = \sqrt{\frac{(2-11/3)^2+(4-11/3)^2+(5-11/3)^2}{3-1}} \approx 1.5275$

$= \sqrt{\frac{3}{3-1}}$, $\sqrt{\overline{x^2}-\bar{x}^2} = \sqrt{3/2}$ $\sqrt{15-3.666^2} \approx 1.5275$

14. $\hat{\sigma}_x = \sqrt{\frac{\sum(x-\bar{x})^2}{n}} = \sqrt{\overline{x^2}-\bar{x}^2} = \sqrt{15-3.6667^2} \approx 1.2472$

15 = REGR



ELLIPTICAL PLOT

TODAY: INTRO TO MULTIPLE LINEAR REGRESSION.

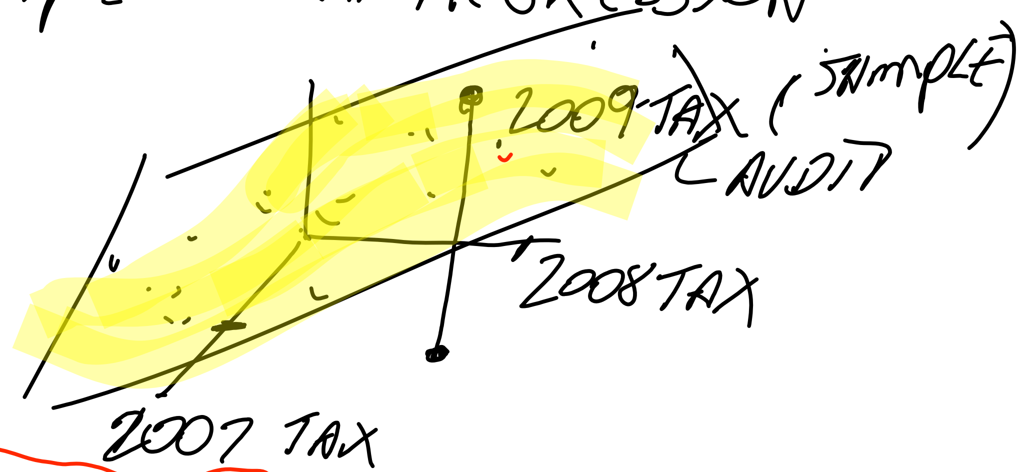
MODEL:

$$y = b_0 + b_1 x_1 + b_2 x_2$$

y: 2009 TX (AUDIT)

x_1: 2008 TX

x_2: 2007 TX



LS (MLR. SAME) $\hat{b}_0 \quad \hat{b}_1 \quad \hat{b}_2$

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \qquad A = \begin{array}{l} \begin{pmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \end{pmatrix} \end{array} \begin{array}{l} \text{FIRST} \\ \text{SECOND} \end{array}$$

betahat[A , $\{y_1 \ldots y_n\}$] → $\hat{b}_0, \hat{b}_1, \hat{b}_2$

y_1 = 2009 AUDIT TX SUBJECT 1

x_{11} 2008 TX SUBJECT 1

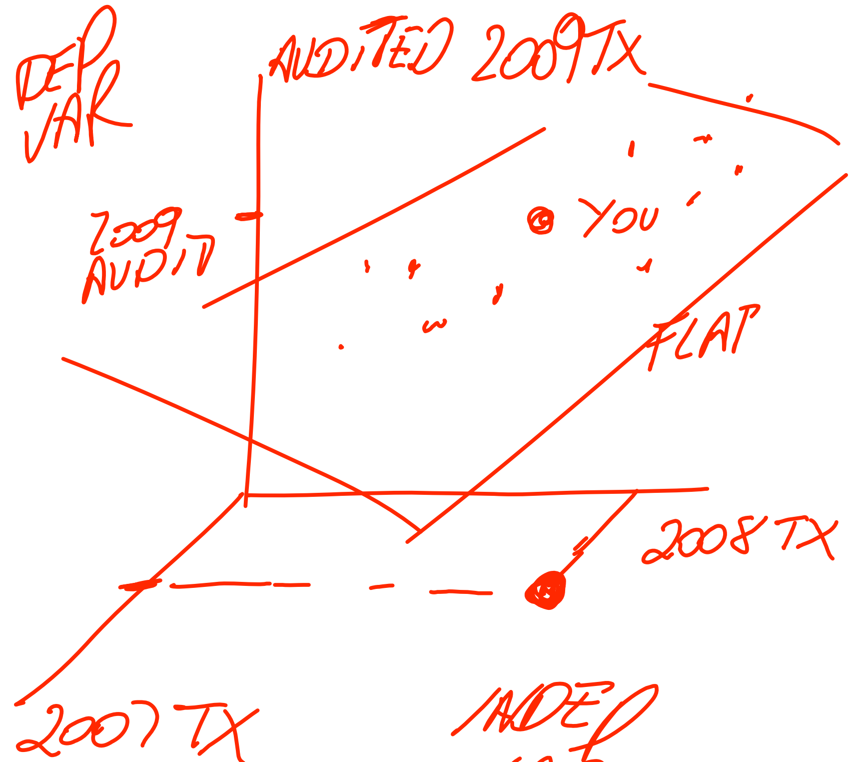
x_{21} 2007 TX

DATA IN → $\hat{b}_0, \hat{b}_1, \hat{b}_2$

R MULTIPLE CORRELATION

$= \rho[\text{FITTED}, y]$

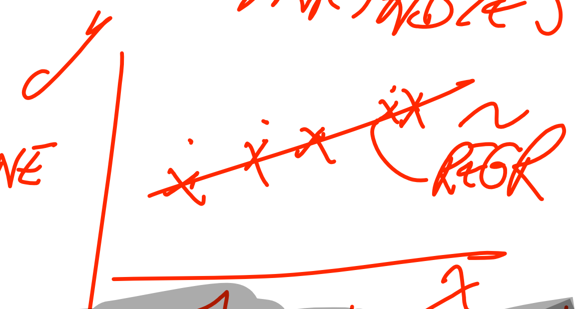
ORDINARY



CLAIM: IN STRAIGHT LINE REGRESSION

$$\rho[\hat{y}, y] = |\rho[x, y]|$$

NOTE ST LINE



CLAIM: $\rho[x, \hat{y}] = |\rho[x, y]|$

TYPO